DOCUMENT RESUME

ED 036 167 24 EM 007 753

AUTHOR LINDVALL, C. M.; NITKO, ANTHONY J.
TITLE CRITERION-REFERENCED TESTING AND THE INDIVIDUALIZATION OF INSTRUCTION.
INSTITUTION RESEARCH FOR BETTER SCHOOLS, INC., PHILADELPHIA, PA.
SPONS AGENCY OFFICE OF EDUCATION (DHEW), WASHINGTON, D.C. BUREAU OF RESEARCH.
BUREAU NO BR-6-2867
PUB DATE 69
CONTRACT OEC-1-7-062867-3053
NOTE 14P.; PAPER PRESENTED AT ANNUAL MEETING OF NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION (LOS ANGELES, CALIFORNIA, FEBRUARY 6, 1969)

EDRS PRICE EDRS PRICE MF-$0.25 HC-$0.80
DESCRIPTORS DIAGNOSTIC TESTS, EQUATED SCORES, INDIVIDUALIZED INSTRUCTION, *PERFORMANCE TESTS, *TEST INTERPRETATION
IDENTIFIERS INDIVIDUALLY PRESCRIBED INSTRUCTION, IPI

ABSTRACT

TWO WAYS OF INTERPRETING RAW TEST DATA ARE NORM-REFERENCING AND CRITERION-REFERENCING; THE FORMER YIELDS INFORMATION BASED ON SOME TYPE OF ORDERING OF THE PERSON ON THE PERSON'S DIMENSION, WHILE THE LATTER TELLS WHETHER OR NOT STUDENTS CAN EXHIBIT A GIVEN PERFORMANCE. CRITERION-REFERENCED INFORMATION CAN BE ORDERED INTO A CRITERION-REFERENCED SCORE, DEPENDING UPON THE EXTENT TO WHICH IT IS POSSIBLE TO ORDER THE ITEMS ON THE CRITERION DIMENSION. THIS METHOD OF HANDLING INFORMATION IS PARTICULARLY USEFUL IN THE UNIT-OBJECTIVE TESTING METHOD OF INDIVIDUALIZING INSTRUCTION EMPLOYED IN INDIVIDUALLY PRESCRIBED INSTRUCTION (IPI) PROJECTS. (SP)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED036167

CRITERION-REFERENCED TESTING AND THE INDIVIDUALIZATION OF INSTRUCTION

C. M. Lindvall and Anthony J. Nitko

Learning Research and Development Center

University of Pittsburgh

EM007 753 (a)

A paper presented at the Annual Meeting of the National Council on Measurement in Education, February 6, 1969, Los Angeles, California.

CRITERION-REFERENCED TESTING AND THE INDIVIDUALIZATION OF INSTRUCTION

C. M. Lindvall and Anthony J. Nitko

Many recent developments in education have served to emphasize the need for tests and other evaluation techniques that provide information concerning the specific competencies that a pupil does or does not possess rather than information as to how he ranks in comparison with other persons comprising some norm group. Persons concerned with this type of problem have frequently suggested the need for criterion-referenced test scores[1] or content-referenced test scores as opposed to norm-referenced scores. The norm-references test score is, of

---

[1]Robert Glaser, "Instructional Technology and the Measurement of Learning Outcomes: Some Questions," American Psychologist, 1963, 18: 519-521.

---

course, exemplified by most scores on standardized tests where pupil performance is reported in terms of percentile ranks, stanines, grade equivalents, and other scores which tell how his performance compared with that of other persons in some norm group but tell very little about the specific things he does or does not know. It is the purpose of this paper to examine the basic difference between norm-referenced and criterion-referenced scores and to present specific examples of the use of criterion-referenced tests in an actual instructional program.

Basic Rationale

One basic task in the evaluation of pupil achievement is that of determining the extent to which a student has achieved certain specific instructional objectives. In its simplest manifestation this involves the determination of whether or not one person can exhibit one specific capability. Can he tie his own shoelaces? Can he pronounce the word cat when he sees it in print? Can ge give the correct answer to 2 plus 3? If we have a yes or no answer for any such question for some one individual we have criterion-referenced information. We know whether or not this person can exhibit some specific performance. It is proposed here that this is the basic element in achievement testing or in any type of evaluation of achievement. That is, the basic element is a yes or no concerning a person's ability to display some specific performance.

To examine how this basic element plays a part in the reporting and analysis of evaluation data it is useful to consider a two-way table in which the marginal entries are persons and types of performances. A simple illustration of this is provided in Table 1. Here the column headings identify test items that measure knowledge of simple addition facts while the row headings are names of specific students. Each cell in the table provides specific, criterion-referenced information for a given student. We might choose to always report our evaluation data in just this form. It is very informative, for example, to be able to report that Jon Smith has command of the facts 1 + 1, 1 + 2, 1 + 3, and 2 + 2 but does not have command of the facts 2 + 3 and 4 + 1. This information is "criterion-referenced" and could be very useful for instructional planning. We can use our information in this way to distinguish among items, i.e., to report which items a person has mastered and which he hasn't. We can

also choose to use the other dimension and distinguish among students. For example, we might report that Bob and Pat have mastered the problem 2 + 3 = 5 while Sue, Jim, and Jon have not. Note that these are still "criterion-referenced" reports. They tell us whether or not students can exhibit a given performance.

In some cases we may choose to combine groups of test items into some larger block. For example, in Table 1 the six items, when combined, may be considered as measuring a pupil's command of the simple addition facts with sums of five or less. If we arbitrarily decide that the student who answers at least five out of the six problems correctly has command of this set of facts, we would arrive at the criterion-referenced decisions indicated in the right hand column of the table. Here we have somewhat less specific information than that provided by individual item data, but it is still criterion-referenced information to the extent that it tells us whether a pupil has or has not mastered some definable domain of tasks. (In this case, knowledge of the simple addition facts with sums of five or less.)

Just as we may choose to combine groups of test items, there may be situations in which we would choose to report information on groups of pupils, such as all students in a class. This is exemplified by the last row in Table 1 where the "yes" is intended to indicate that the class has mastery of a given item and where this decision is based on whether or not at lease 80 per cent of the class showed mastery of it. How we set up this table and how we combine or do not combine cells are the determiners of what kind of information we get from the table or from our test.

It should be obvious from the foregoing discussion that criterion-referenced test information is here defined as the type of information

that tells us that a person (or a group) can exhibit these specific performances and/or cannot exhibit these specific performances. Below we will explain how such information can, under certain conditions, be presented in the form of criterion-referenced scores, but it is essential to realize that the use of tests to achieve criterion-referenced information is not dependent upon the possibility of deriving such scores. It is dependent only upon the possibility of being able to describe what a person can and cannot do.

Deriving Scores

Test scores are based on some type of count of the number of items answered correctly by a student. Such raw scores have limited meaning. There are two basic approaches that may be followed in the attempt to give more meaning to such a score. One is to attempt to give the score a criterion-referenced meaning. The other is to give it a norm-referenced meaning.

Norm-Referenced Scores

Typical norm-referenced scores include percentile ranks, age equivalents, grade equivalents, stanines, and standard scores. To return to our conceptualization of test results as being based on data such as that presented in Table 1, it can be said that norm-referenced scores are based on some type of ordering of the persons on the persons' dimension. This is exemplified by the simple example in Table 2 which can be considered as being derived from a table, such as Table 1, where we have added up the total number of items correct over at least 53 items and have then rearranged the persons in our row headings so that they are in descending order according to the magnitude of their total scores.

Criterion-Referenced Scores

Previously we have pointed out that we obtain criterion-referenced information from a test by describing exactly what items a person is able to answer correctly and what items he cannot answer correctly. Transforming such possibly lengthy descriptive information into a criterion-referenced score can be shown to be dependent upon the extent to which it is possible to order the items on the criterion dimension (just as the derivation of norm-referenced scores is dependent upon the ordering of the entries on the person dimension). To illustrate this point, consider the type of data presented in Table 3. It is probably not hard to imagine that results such as those shown might be obtained when this six-item test was given to students of the appropriate grade level. Note here that persons can get the same score even though they have mastered different combinations of addition facts. Now let us picture another test such as that presented in Table 4. In this situation the items on the test appear to represent noticeable differences in the prerequisite nature of the learning which they attempt to measure and it is possible to order the items to reflect this prerequisite learning sequence.

The derivation of a criterion-referenced score would seem to demand results that, quite consistently, followed the type of pattern shown in this figure. In this case all persons with a score of 3 have answered the same three items correctly. The same is true for a score of 2, and presumably would be true for a score of 1. Knowledge of a person's score tells you exactly what things he is able to do and what things he is not able to do. The score is a criterion-referenced score. (What we are picturing here is a set of scores having perfect scalability and

reproducibility in the Guttman sense.[2]) Note that the derivation of

---

[2]Louis Guttman, "A Basis for Scaling Qualitative Ideas," American Sociological Review, 1944, 9:139-150.

---

such a criterion-referenced score is dependent upon the possibility of ordering items in a sequence that consistently manifests itself in the way in which persons perform on the test. In actual applications of this procedure it may be necessary to score this type of test in terms of units made up of groups of items rather than single items. For example, if we had a sequenced scale of related performances (such as increasing competencies in addition) made up of fifty such performances and the related test items, we might find it necessary and useful to divide this into ten groups of five items each. Such a test would be constructed so that each five-item group of the test would be described as measuring some one domain of performances, and the pupil would be scored 1 or 0 (pass or fail) on each group. It is this quality of being able to order the items or groups of items on a test, and to have this order consistently validated by the way in which students actually answer items, which seems to be essential to the derivation of criterion-referenced scores.

Since this is the case, it may well be that criterion-referenced scores of the type being proposed here cannot be used to make fine discriminations as to where a pupil is located on some continuum. It may be necessary to use such scores to locate pupils with respect to the criterion scale only on a relatively gross basis and then to use some type of item-by-item analysis to obtain more specific data.

However, it should be remembered that criterion-referenced information as previously defined is the real need and that criterion-referenced scores are merely a more convenient and efficient way of handling such information. Criterion-referenced information can be obtained in any situation where one is willing to take the time to spell out performance objectives, to develop items and tests to assess each objective, and to examine results in whatever way is necessary for gaining the required information.

An Example of the Use of Criterion-Referenced Testing

One practical example of the problems encountered in the application of the rationale developed in this paper is found in the testing program used with Individually Prescribed Instruction.[3] IPI is a procedure for

---

[3]C. M. Lindvall and John O. Bolvin, "Programmed Instruction in the Schools: An Application of Programming Principles in Individually Prescribed Instruction," Programmed Instruction, Sixty-Sixth Yearbook of the National Society for the Study of Education, Part II, (Chicago, Illinois: University of Chicago Press, 1967).

---

individualizing instruction in the elementary school and involves the specification of sequences of units and of objectives, the development of tests to measure pupil performance on each objective and each unit, and the use of procedures that permit each pupil to start at his appropriate point in the curriculum and to proceed at his own individual pace. Some idea of how criterion-referenced tests may be employed in individualized instruction can be obtained by examining the procedure used in IPI for starting each student at his appropriate point in the curriculum.[4]

---

[4]C. M. Lindvall and Richard C. Cox, "The Role of Evaluation in Programs for Individualized Instruction," Yearbook of the National Society for the Study of Education, (Chicago, Illinois: University of Chicago Press, 1969).

The way in which the IPI math curriculum is structured may be seen by examining Table 5. It will be noted that the curriculum is organized in terms of topics (Numeration, Place Value, etc.) and levels (Level A, Level B, etc.). A given topic at a given level, such as Level B Addition, constitutes a unit, and each unit involves some number of specific objectives. Getting a student started at the proper point in the math curriculum involves determining the unit in which he should start and also which objectives in that unit he should study. The IPI testing program has been developed to accomplish this. It will be noted from Table 5 that IPI math is organized in terms of relatively homogeneous topics that are studied at progressive levels of difficulty as the student works from Level A, to Level B, to Level C, and so on. The first task of placement testing then is to determine to what level a student has progressed in each of these topics. The topics have been developed in a way to make the progression from level to level represent a prerequisite hierarchy in which the abilities learned at each level build on those acquired at the preceding level and are prerequisite to those to be learned at the next level. In this sense, the sequence of levels within each topic (for example, A Numeration, B Numeration, C Numeration, D Numeration, etc.) constitute a hierarchy. Placement testing first involves finding where the student's capabilities place him along this hierarchy. For example, placement testing within the Numeration topic involves determining that a student has mastered levels A, B, C, and D but has not mastered any levels above this. In essence, the report is that he has a "score" of level D in Numeration. Note that this is a criterion-referenced score.

Because of the relatively gross nature of the information provided by these placement test scores, further criterion-referenced testing must be employed before a student actually starts instruction in any topic.

The score of "level D" in Numeration obtained by the hypothetical student described above tells us that he is ready to start work in Level E in the Numeration continuum. However it is also important to determine whether or not he has mastered any of the specific performances identified by the six objectives in Level E. That is, his placement test score tells us that he has not mastered all of level E but this does not preclude the possibility that he has mastered some of the individual objectives. To determine whether or not this is the case, we need additional criterion-referenced information.

If the objectives in Level E Numeration could be sequenced in a prerequisite order, it should be possible to develop a scaled test yielding a criterion-referenced score.[5] Up to this time IPI unit tests are only

---

[5]For an example suggesting the possibility of doing this see Richard C. Cox and Glen T. Graham, op. cit.

---

rough approximations to this ideal and do not yield such scaled scores. However the IPI program does employ criterion-referenced tests which yield criterion-referenced information at this point. Such tests are known as the unit pretests. These tests are structured so as to provide a sub-score for each objective within the unit and are scored so as to indicate whether the student has mastered or has not mastered each objective. This criterion-referenced information tells the teacher what the pupil can and cannot do with respect to the skills covered in this unit and enables him to make instructional decisions concerning what the pupil should study. Thus, a combination of criterion-referenced scores from the placement tests and criterion-referenced information from the unit test serves to provide rather exact information concerning the specific competencies that the pupil does and does not possess.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

CRITERION-REFERENCED TESTING AND THE INDIVIDUALIZATION OF INSTRUCTION

C. M. Lindvall and Anthony J. Nitko

Learning Research and Development Center

University of Pittsburgh

EM007 753(6)

Handouts for a paper presented at the Annual Meeting of the National Council on Measurement in Education, February 6, 1969, Los Angeles, California.

Table 1

Basic Information Necessary for Developing Summary Data Regarding Test Performance: Command (yes) or Lack of Command (no) of Specified Criterion Performance by Individual Student

Persons	Test Item (Addition Fact)						Total Test
	1 +1	1 +2	1 +3	2 +2	2 +3	4 +1	
Bob Adams	yes	yes	yes	yes	yes	no	yes
Sue Bond	yes	yes	yes	yes	no	yes	yes
Jim Carr	yes	no	yes	no	no	no	no
. . . .							
. . . .							
Jon Smith	yes	yes	yes	yes	no	no	yes
. . . .							
. . . .							
Pat Yates	yes	yes	yes	yes	yes	yes	yes
Total Class	yes	yes	yes	yes	no	no	yes

Table 2

Raw Scores and Percentile Ranks for Ten Persons Arranged in Order of Size

Person	Raw Score	Percentile Rank
Rose	53	95
Paul	47	85
Alma	46	75
Pat	43	65
Terry	40	55
Alex	38	45
Dianne	33	35
Mary	31	23
Art	28	15
Tony	26	5

Table 3

Possible Results for Students Taking Six-Item Tests on Addition Facts

Persons	Test Item (Addition Fact)						Total Score
	1 +1	1 +2	1 +3	2 +2	2 +3	4 +1	
Jack	1	1	1	1	0	0	4
Ray	1	1	0	1	1	0	4
Mae	1	1	0	1	0	0	3
Ann	1	0	1	0	0	1	3

Table 4

Possible Results for Students Taking Four-Item Addition Test: Items Ordered by Increasing Difficulty

Persons	Test Item				Total Score
	2 +1	6 +7	23 +65	87 +69	
Sue	1	1	1	1	4
Randy	1	1	1	0	3
Dick	1	1	1	0	3
Bill	1	1	0	0	2
Ruth	1	1	0	0	2
Art	1	0	0	0	1

Table 5

Number of Instructional Objectives at Each Level for Each Topic (or in Each Unit in the IPI Mathematics Curriculum)

Topic	Level							
	A	B	C	D	E	F	G	H
Numeration	9	7	3	3	6	3	6	6
Place Value			2	4	3	5	1	1
Addition	2	9	5	8	6	2	4	3
Sub.			3	5	3	1	3	1
Mult.				8	10	10	4	3
Div.				7	7	5	5	5
Comb.			4	5	6	4	5	5
Fractions	2	1	4	5	6	12	7	1
Money		4	2	5	4	1		
Time		3	5	10	16	5		
Systems		4	2	4	6	2		
Geom.		2	2	1	9	9	6	6
Spec.				3	3	5	3	3
Supp. Topics			3	1	1	1	1	1